Final Exam

601.467/667 Introduction to Human Language Technology

Fall 2024
Johns Hopkins University
Co-ordinator: Philipp Koehn

17 December 2024

Complete all questions.

Use additional paper if needed.

Time: 75 minutes.

Name of student: _____

Q1. Question Answering *20 points*

1. Methods from several subfields of NLP (namely traditional Question Answering, Machine Reading Comprehension, and Large Language Models) can all address the task of answering questions. For example, I can ask the same question "What countries are the largest producers of lithium?" to IBM Watson, a SQuAD model, and ChatGPT. Please discuss the ways in which these three subfields differ, in terms of motivation and problem setup. (12 points)

2. Continuing from the previous question: if you were to build a start-up centered around answering questions about research papers to sell to a grad student audience, which of the three approaches (QA, MRC, LLM) would you implement? Pick one and justify. (3 points)

3. What are the five main components of a Question Answering system? Draw a flow chart that explains how these components combine, and how an input question leads to an output answer. (5 points)

Q2. Digital Humanities *20 points*

1. Human and computational intelligence differ in a number of ways. Describe a way in which they are *complementary*, and give a pair of examples illustrating when a human would offer greater insight than a computer, and vice versa (10 points)

2. Graph convolutions are a simple generalization of the familiar ”grid” convolutions used throughout HLT and computer vision. Explain this generalization, specifically addressing how the ”receptive field” grows in the two architectures (10 points)

Q3. Human-Centered Evaluation *20 points*

1. Imagine you are evaluating a new summarization. You need human-annotated references for evaluating its performance. One option is to hire professional linguists to create the annotations, while another is to use a crowdsourcing platform like Amazon Mechanical Turk. What are the trade-offs of these two methods? And what are the potential issues of using reference-based method evaluation for such a system?

2. Explain the role of reliability and validity in the context of language technology evaluation.

3. What's the difference between empirical and analytical evaluation methods for language technology?

4. What's the difference between the normative benchmark and user study regarding human requirement realism, context realism, and pragmatic cost?

Q4. Ethical Problems *20 points*

1. List four ethical principles in AI and briefly explain them (5 points).

2. Explain what an Institutional Review Board is and its role to ensure an ethical AI (5 points).

3. How does HIPAA (Health Insurance Portability and Accountability Act) try to protect citizen's rights? (5 points)

4. How can we measure (and avoid) maleficence or unfairness in AI? (5 points)

Q5. Computational Social Science *20 points*

1. What are two ways pre-trained language models like BERT or RoBERTa have been used for computational social science research? Provide both a broad description of the methodology and a specific example of how this method is useful for a particular research question, as in the provided example (8 points)

   Example: Prompting BERT-style models with template sentences can be used for metaphor detection. Metaphor detection can be used to identify dehumanizing language in political speeches and how it has varied over time.

   1.

   2.

2. What are two ways GPT-style models might be useful for computational social science research? As before, provide both a broad description of the methodology and a specific example of how this method is useful for a particular research question (8 points)

   1.

   2.

3. Label the following statements about topic models as true or false (4 points):

   In LDA, "topics" are defined as distributions over the vocabulary _____

   The goal of LDA is to estimate K, the number of topics in the corpus _____

   Topic models require in-domain annotated data _____

   Topic models are useful for datasets where researchers have specific well-defined research questions _____